DOCUMENT RESUME

ED 315 429 TM 014 444

AUTHOR	Gardner, Eric
TITLE	Five Common Misuses of Tests. ERIC Digest No. 108.
INSTITUTION	American Institutes for Research, Washington, DC.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Mar 89
CONTRACT	RI-88062003
NOTE	3p.; Reprinted from "Ability Testing: Uses, Consequences, and Controversies," 1982.
PUB TYPE	Information Analyses - ERIC Information Analysis Products (071) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Error of Measurement; Evaluation Problems; Examiners; Scoring; *Statistical Analysis; *Testing Problems; Test Interpretation; *Test Use
IDENTIFIERS	ERIC Digests

ABSTRACT

Five of the common misuses of tests are reviewed: (1) acceptance of the test title as an accurate and complete description of the variable being measured (failure to examine the manual and the items carefully to know the specific aspects to be tested can result in misuse through selection of an inappropriate test for a particular purpose or situation); (2) ignoring the error of measurement in test scores; (3) use of a single test score for decision making (scores are not interpreted in the full context of the various elements that characterize students, teachers, and the environment); (4) a lack of understanding of the meaning of test score reporting (the misinterpretation of raw scores or grade equivalents is common); and (5) attributing cause of behavior measured to test (confusing the information provided by a test score with interpretations of what caused the behavior or described by the score). A test score gives no information as to why the individual performed as reported. No statistical manipulation of test data will permit more than probabilistic inferences about causation or future performance. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ERIC Clearinghouse on Tests, Measurement, and Evaluation Digest No. 108, March 1989

ED315429

Five Common Misuses of Tests

by Eric Gardner

Reprinted from "Ability Testing: Uses, Consequences, and Controversies," 1982, with permission from the National Academy Press, Washington, DC

Acceptance of a Test Title for What the Test Measures

There is a tendency for unsophisticated test users to accept the name assigned to a test as an accurate and complete description of the variable being measured. Since titles must be brief, they cannot convey all that the user needs to know about the kind of behavior to be measured. All tests are open to this kind of uncritical abuse. Since there are so many facets of cognitive ability, it is obvious that no test can be an adequate measure of them all. Only full knowledge of the items can reveal what is being measured. Furthermore, the testing situation may completely change the expected behavior.

If a non-English speaking or blind pupil is given an "aptitude" test in printed English, it obviously doesn't measure any aspect of "aptitude" or "intelligence" except lack of knowledge of English or lack of vision. In a less obvious area, a test labeled "Science Achievement" may be an acceptable test to sample the science curriculum for students in a particular fifth grade science course but fail to function as a science test at all for most pupils if the reading difficulty is at the high school level. A test producer's claims for an achievement test or an aptitude test do not mean that it will function as such in all circumstances with all pupils. Failure to examine the manual and the items carefully in order to know the specific aspects of cognitive ability to be tested (memory, vocabulary, type of reasoning, etc.) can result in misuse by virtue of selecting an inappropriate test for a particular purpose or situation.

Ignoring the Error of Measurement in Test Scores

Every test score contains an error of measurement. It is a misuse of any test score or any observation to accept the test score as a fixed, unchanging index containing no error. It is impossible to say with certainty that an individual's observed score gives his or her "true" performance on the general domain about which inferences are to be made. The best that can be done is to estimate experimentally the standard error of measurement and then use that value to set up a band within which a probability can be stated about the "true" score's being within that band. That

- we cannot accept an SAT score of 550 as a precise measure,
- we must accept a range of scores, and
- we must then expect to be wrong a certain proportion of the times

does not mean that the SAT does not furnish useful data. It does mean that the test score is being misused if knowledge of the size of the errors of measurement is not used in interpreting the score.

In the case of most standardized test scores, the magnitude of the errors is made explicit, not hidden or unknown. In fact, the errors in essay grading or any other type of evaluative data have far larger but usually unknown errors of measurement.

Some people reject the notion of basing decisions on probabilistic data. However, probability estimates are involved in almost all decisions. For example, the decision to cross a busy street at a particular instant is not made with a probability of 1.0 of doing so safely.

Use of a Single Test Score for Decision Making

Misuse of tests occurs when scores are not considered and interpreted in the full context of the various elements that characterize pupils, teachers, and the general educational environment involved because a test score represents only a sample from a limited

TM014444

domain and does not include the variety of factors that might influence that score. For example, in decisions determining admission to college, SAT scores should not be used in isolation and are, in fact, usually considered along with the pupil's high school record and other relevant data, such as teacher's or supervisor's recommendations concerning motivation, leadership ability, creativity, involvement in extracurricular activities, etc.

All of these can then be evaluated against the student's socioeconomic background, along with consideration of any social obstacles or unusual physical demands required of the student to reach his or her current educational level.

Lack of Understanding of Test Score Reporting

There is substantial misunderstanding, not just among laymen, but also among many educators, of the meaning of test scores. Most people believe that they understand the meaning of a raw score or of that particular raw score converted to a percent of items answered correctly, as in the case of many criterion-referenced tests. However, even in this most elementary illustration;, more is involved than a single number indicates. Forty-five items answered correctly out of 50 easy items has a substantially different meaning than 45 items answered correctly out of a sample of 50 very difficult items from the same domain.

The interpretation of a raw score converted to a percentile score causes even more problems. The statement that "In a norm-referenced test, half the pupils must fail" is a good nor a poor performance. It merely indicates that among the group used as a frame of reference this score was higher than that reached by 20% of its members. If the group were of high ability or had unusual skills, a percentile rank of 20 might indicate an excellent or even remarkable performance.

The misinterpretation of grade equivalents is even more common. A grade equivalent is the score that was exceeded by 50% of the group at the specific time when the test was given. It does not represent a standard to be attained. It does not represent the grade in which the pupil should be placed.

To compensate for the decreasing emphasis on test construction and test interpretation in teacher training institutions, there have been efforts by the National Council on Measurement in Education[1] and other organizations to provide workshops and reading material on measurement issues. Both parents and professional educators stand to benefit, since both are involved in the misuse of testing based on misinterpretation of scores.

Attributing Cause of Behavior Measured to Test

It is common, especially for critics of testing, to confuse the information provided by a test score with interpretations of what caused the behavior described by the score. A test score is a numerical description of a sample of performance at a given point in time. A test score gives no information as to why the individual performed as reported.

Claiming that it does, whether intended as a positive attribute or a criticism, is tantamount to test misuse. Furthermore, no statistical manipulation of test data, even though combined with the best additional data, will permit more than probabilistic inferences about causation or future performance.

The reports on the decline of SAT scores is an excellent example of the difficulty in ascribing causation to known performance. The charge given the researchers by the investigating panel was to explain the causes of the drop in SAT scores. They were able to describe the drop and offer changes in test populations as a plausible partial explanation for the initial drop but could only speculate on the effect of other variables and the reasons for the continued drop.

1 The National Council on Measurement in Education (NCME) is a national organization of professionals concerned with testing and measurement issues.

Additional Reading (compiled by ERIC/TM)

Woodring, Paul (Dec 16, 1987) Irresponsible News Stories on SAT Scores Misuse the Facts and Lead to Confusion. *Chronicle of Higher Education*, 34, 16, B1.

Kearney, C. Philip (Fall, 1983) Uses and Abuses of Assessment and Evaluation Data by Policymakers. *Educational Measurement: Issues and Practice*; 2 3, 9-12.

Green, Donald Ross (1985) *Misinterpreting and Misusing Tests: Some New Ways.* ERIC Document Reproduction Service, ED 291 805.

Echternacht, Gary (1981) *The Uses and Misuses of Test Scores:* Technical Assistance Perspective Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA, April 13-17, 1981 (ED 199 275).

This publication was prepared with funding from the Office of Educational Research and Improvement (OERI), U.S. Department of Education, under contract R-88-062003. The opinions expressed in this report do not necessarily reflect the position or policy of OERI or the Department of Education. **Permission is granted to copy and distribute this ERIC/TM digest.**